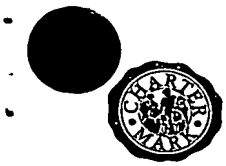


INVESTOR IN PEOPLE

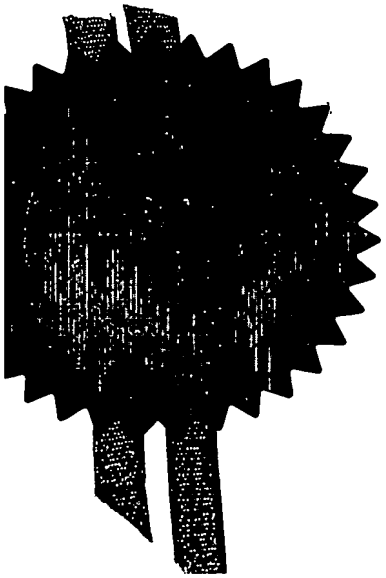
The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

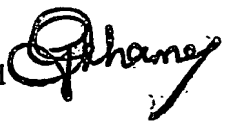
I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed 

Dated 17 February 2005

BEST AVAILABLE COPY

Patents Form 1/77

Pate[illegible]ct 1977
(Rule 16)

THE PATENT OFFICE
SW
10 NOV 2003
LONDON
The Patent Office
PATENTS • DESIGNS • COPYRIGHT • TRADE MARKS

1/77

11NOV03 E851050-11 D00016
P01/7700 0.00-0326201.1

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

1. Your reference — NS/03497GB

2. Patent application number *(The Patent Office will fill in this part)* — 0326201.1 — 10 NOV 2003

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

Forbidden Technologies PLC
2-4 St. George's Road
Wimbledon
LONDON SW19 4DP

Patents ADP number *(if you know it)* 7936313001

If the applicant is a corporate body, give the country/state of its incorporation — United Kingdom

4. Title of the invention — Improvements Relating to Representations of Compressed Video

5. Name of your agent *(if you have one)* — BROOKES BATCHELLOR LLP

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

102-108 Clerkenwell Road
London EC1M 5SA

Patents ADP number *(if you know it)* 08839276001 ~~08142291001~~

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

Yes

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	19
Claim(s)	0
Abstract	0
Drawing(s)	5 + 5 Sht

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Brooker Bakewell Date 10/11/2003

12. Name and daytime telephone number of person to contact in the United Kingdom Nigel Shindler 020 7253 1563

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

Forbidden Technologies plc, 10 November 2003

DUPLICATE

Improvements Relating to Representations of Compressed Video

This invention relates to a method of processing of digital video information. This digital video information is either compressed for storage and then later transmission, or is compressed and transmitted live with a small latency. Transmission is for example over the internet.

There is a need for highly efficient compression techniques to be developed to enable transmission of video in real time over the internet because of the restrictions in the bandwidth. In addition, the increasing need for high volume of content and rising end-user expectations mean that a market is developing for live compression at high frame rate and image size.

An object of the invention is to provide such compression techniques.

General background

The video to be compressed can be considered as consisting of a number of frames (at least 1), each made up of individual picture elements, or pixels. Each pixel can be represented by three components, usually either RGB (red, green and blue) or YUV (luminance and two chrominance values). These components can be any number of bits each, but eight bits of each is usually considered sufficient.

these regions is represented by one UV pair.

Real time filtering.

The aim is to remove noise from the input video, as noise is by definition hard to compress. The filtering mechanism takes frames one at a time. It compares the current frame with the previous filtered frame on a pixel-by-pixel basis. The value for the previous pixel is used unless there is a significant difference. This can occur in a variety of ways. In one, the value of the pixel in the latest frame is a long way from the value in the previous filtered frame. In another, the difference is smaller, but consistently in the same direction. In another, the difference is even smaller, but cumulatively, over a period of time, has tended to be in the same direction. In these the first two cases, the pixel value is updated to the new value. In the third case, the filtered pixel value is updated by a small amount in the direction of the captured video. The allowable error near a spatial edge is increased depending on the local contrast to cut out the effects of spatial jitter on the input video.

Real time motion estimation

The video frames are filtered into "Noah regions". Thus the pixels near to edges are all labelled. In a typical scene, only between 2% and 20% of the pixels in the image turn out to have the edge labelling. There are three types of motion estimation used. In the first, whole frame pan detection using integer number of pixels is implemented. These motions can be implemented efficiently over the whole image

on playback as pixels can be copied to new locations and no blurring is needed. This uses the edge areas from the Noah regions only as the edges contain the information needed for an accurate motion search. The second is sub-pixel motion removal over the whole image. This uses the edge areas from the Noah regions only as the edges contain the information needed for an accurate motion search. The edge pixels in the image, estimated by example from the Noah filtering stage, are matched with copies of themselves with translations of up to eg 2 pixels, but accurate to eg 1/64 pixel (using a blurring function to smooth the error function) and small rotations. The best match is calculated by a directed search starting at a large scale and increasing the resolution until the required sub-pixel accuracy is attained. This transformation is then applied in reverse to the new image frame and filtering continues as before. These changes are typically ignored on playback. The effect is to remove artefacts caused by camera shake, significantly reducing datarate and giving an increase in image quality. The third ~~for~~ examines local areas of the image. Where a significant proportion of the pixels are updated, for example on an 8x8 pixel block, either motion vectors are tested in this area with patches for the now smaller temporal deltas, or a simplified super-block representation is used giving either 1 or 2 YUVs per block, and patches are made to this.

Real time fade representation

The encoding is principally achieved by representing the differences between consecutive compressed frames. In some cases, the changes in brightness are

spatially correlated. In this case, the image is split into blocks or regions, and codewords are used to specify a change over the entire region, with differences with these new values rather than differences to the previous frame itself being used.

Segment Noah regions - find edges

A typical image consists of areas with low contrast and areas of high contrast, or edges. The segmentation stage described here analyses the image and decides whether any pixel is near an edge or not. It does this by looking at the variance in a small area containing the pixel. For speed, in the current implementation, this involves looking at a 3x3 square of pixels with the current pixel at the centre, although future implementations on faster machines can be expected to look at a larger area. The pixels which are not near edges are compressed using an efficient but simple representation which includes multiple pixels - for example 2x2 blocks or 8x8 blocks, which are interpolated on playback. The remaining pixels near edges are represented as either eg 8x8 blocks with a number of YUV areas (typically 2 or 3) if the edge is simply the boundary between two or more large regions which just happen to meet here, or as 2x2 blocks with 1Y and one UV per block in the case that the above simple model does not apply eg when there is too much detail in the area because the objects in this area are too small.

Miniblockify

The image is made up of regions, which are created from the Noah regions. The

relatively smooth areas are represented by spatially relatively sparse YUV values, with the more detailed regions such as the Noah edges being represented by 2x2 blocks which are either uniform YUV, or consist of a UV for the block and maximum Y and a minimum Y, with a codeword to specify which of the pixels in the block should be the maximum Y value and which should be the minimum. To further reduce the datarate, the Y pairs in the non-uniform blocks are restricted to a subset of all possible Y pairs which is more sparse when the Y values are far apart.

Transitions with variable lengths codewords

Compressing video consists in part of predicting what the next frame will as accurately as possible from the available data, or context. Then the (small) unpredictable element is what is sent in the bitstream, and this is combined with the prediction to give the result. The transition methods described here are designed to facilitate this process. On compression, the available context and codeword to compress are passed to the system. This then adds this information to its current distribution (which it is found performs well when it starts with no prejudice as the the likely relationship between the context and the output codeword). The distribution output data for this context is calculated and variable length codewords assigned to the outcomes which have arisen. These variable length codewords are not calculated each time the system is queried as the cost/reward ratio makes it unviable, particularly as the codewords have to be recalculated on the player at the corresponding times they are calculated on the compressor. Instead, the codewords

are recalculated from time to time. For example, every new frame, or every time the number of codewords has doubled. Recalculation every time an output word is entered for the first time is too costly in many cases, but this is aided by not using all the codeword space every time the codewords are recalculated. Codeword space at the long end is left available, and when new codewords are needed then next one is taken. As these codewords have never occurred up to this point, they are assumed to be rare, and so giving them long codewords is not a significant hindrance. When the codeword space is all used up, the codewords are recalculated. The minimum datarate for Huffman codewords is a very flat and wide minimum, so using the distribution from the codewords which have occurred so far is a good approximation to the optimal. Recalculating the codewords has to happen quickly in a real time system. The codewords are kept sorted in order of frequency, with the most frequent codewords first. The sorting is a mixture of bin sort using linked lists which is O(n) for the rare codewords which change order quite a lot, and bubble sort for the common codewords which by their nature to not change order by very much each time a new codeword is added. The codewords are calculated by keeping a record of the unused codeword space, and the proportion of the total remaining codewords the next data to encode takes. The shorted codeword when the new codeword does not exceed its correct proportion of the available codeword space is used. There is are further constraints - in order to keep the codes as prefix codes and to allow spare space for new codewords, codewords never get shorter in length, and each codeword takes up an integer power of two of the total codeword space. This method creates the new codewords into a lookup table for quick encoding in

O(n) where n is the number of sorted codewords.

Memory management

To facilitate Java playback, all the memory used is allocated in one block at the start. As garbage collection algorithms on Java virtual machines are unpredictable, and many stop the system for periods which are long in comparison to the length of a video frame, the invention may use its own memory management system. This involves allocating enough memory for eg 2 destination codewords for each source codeword when it is first encountered. New transitions are added as and when they occur, and when the available space for them overflows, the old memory is ignored, and new memory of twice the size is allocated. Although up to half the memory may end up unused, the many rare transitions take almost no memory, and the system scales very well and makes no assumptions about the distribution of transitions.

Give compressed codeword for this uncompressed codeword

Every time a codeword occurs in a transition for the second or subsequent time, its frequency is updated and it is re-sorted. When it occurs for the first time in this transition however, it must be defined. As many codewords occur multiple times in different transitions, the destination value is encoded as a variable length codeword each time it is used for the first time, and this variable length codeword is what is sent in the bitstream, preceded by a "new local codeword" header codeword.

Similarly, when it occurs for the first time ever, it is encoded raw preceded by a "new global codeword" header codeword. These header codewords themselves are variable length and recalculated regularly, so they start off short as most codewords are new when a new environment is encountered, and they gradually lengthen as the transitions and concepts being encoded have been encountered before.

Video compression (cuts)

Cuts are compressed using spatial context from the same frame.

Cuts, RLE uniform shape, else assume independent and context=CUT_CW.

Cuts -> editable, so needs efficient. First approximation at lower resolution eg 8x8.

Cuts - predict difference in mini-block codewords from previous one and uniform flag for current one.

Video compression (deltas)

The deltas can use temporal and spatial context.

Deltas shape - predict shape from uniformness of four neighbours and old shape.

Deltas - predict mini-block codeword differences from uniformness of this mini-block and old mini-block in time.

Datarate reductions

Various simple but effective datarate reduction methods are employed. Noise in the input signal can lead to isolated small changes over the image, whose loss would not be noticed. Isolated changed mini-blocks are generally left out from the bitstream, though if the mini-block is sufficiently different they can still be updated. In addition, small changes in colour in high colour areas are generally ignored as these are almost always caused by noise.

Multi-level gap masks: 4x4, 16x16, 64x64

The bulk of the images are represented mbs and gaps between them. The gaps are spatially and temporally correlated. The spatial correlation is catered for by dividing the image into 4x4 blocks of mbs, representing 64 pixels each, with one bit per mini-block representing whether the mbs has changed on this frame. These 4x4 blocks are grouped into 4x4 blocks of these, with a set bit if any of the mbs it represents have changed. Similarly, these are grouped into 4x4 blocks, representing 128x128 pixels, which a set bit if any of the pixels has changed in the compressed representation. It turns out that trying to predict 16 bits at a time is too ambitious as the system does not have time to learn the correct distributions in a video of typical length. Predicting the masks 4x2 pixels at a time works well. The context for this is the corresponding gap masks from the two previous frames. The transition infrastructure above then works its magic and gives efficient codewords for the gaps at various scales.

Multiple datarates at once

One of the features of internet or intranet video distribution is that the audience can have a wide range of receiving and decoding equipment. In particular the connection speed may vary widely. In a system such as this designed for transmission across the internet, it helps to support multiple datarates. So the compression filters the image once, then resamples it to the appropriate sizes involving for example cropping so that averaging pixels to make the final image the correct size involves averaging pixels in rectangular blocks of fixed size. There is a sophisticated datarate targeting system which skips frames independently for each output bitstream. The compression is sufficiently fast on a typical modern PC of this time to create modem or midband videos with multiple target datarates. The video is split into files for easy access, and these files may typically be 10 seconds long, and may start with a key frame. The player can detect whether its pre-load is ahead or behind target and load the next chunk at either lower or higher datarate to make use of the available bandwidth. This is particularly important if the serving is from a limited system where multiple simultaneous viewers may wish to access the video at the same time, so the limit to transmission speed is caused by the server rather than the receiver. The small files will cache well on a typical internet setup, reducing server load if viewers are watching the video from the same ISP, office, or even the same computer at different times.

Key frames

The video may be split into a number of files to allow easy access to parts of the video which are not the beginning. In these case, the files may start with a key frame. A key frame contains all information required to start decompressing the bitstream from this point, including a cut-style video frame and information about the status of the transition tables, such as starting with completely blank tables.

Digital Rights Management

DRM is an increasingly important component of a video solution, particularly now content is so readily accessible of the internet. Data typically included in DRM may an expiry data for the video, a restricted set of URLs the video can be played from. Once the compressor itself is sold, the same video may be compressed twice with different DRM data in an attempt to crack the DRM by looking at the difference between the two files. The compression described here is designed to allow small changes to the initial state of the transition or global compression tables to effectively randomise the bitstream. By randomizing a few bits each time a video is compressed, the entire bitstream is randomized each time the video is compressed, making it much harder to detect differences in compressed data caused by changes to the information encoded in DRM.

Miscellaneous

The Y values for each pixel within a single super-block can also be approximated. In many cases, there is only one or part of one object in a super-block. In these cases, a single Y value is often sufficient to approximate the entire super-block's pixel Y values, particularly when the context of neighbouring super-blocks is used to help reconstruct the image on decompression.

In many further cases, there are only two or parts of two objects in a super-block. In these cases, a pair of Y values is often sufficient to approximate the entire super-block's Y values, particularly when the context of neighbouring super-blocks is used to help reconstruct the image on decompression. In the cases where there are two Y values, a mask is used to show which of the two Y values is to be used for each pixel when reconstructing the original super-block. These masks can be compressed in a variety of ways, depending on their content, as it turns out that the distribution of masks is very skewed. In addition, masks often change by small amounts between frames, allowing the differences between masks on different frames to be compressed efficiently.

Improvements to image quality can be obtained by allowing masks with more than two Y values, although this increases the amount of information needed to specify which Y value to use.

Brief Descriptions of the Drawings

Figure 1 shows a typical image of 376x280 pixels divided into 8x8 pixel super-blocks.

Figure 2 shows a typical super-block of 8x8 pixels divided into 64 pixels.

Figure 3 shows a typical mini-block of 2x2 pixels divided into 4 pixels.

Figure 4 shows an example image containing two Noah regions and a Noah edge.

Figure 5 shows an example of global accessible context for transition tables.

Figure 6 shows an example of transition tables with local context (LC1 etc) and corresponding resulting values which have been predicted so far

Figure 7 shows the typical context information for cuts.

Figure 8 shows the typical context information for delta frames.

Figure 9 is a flowchart showing how variable length codewords are generated from a list of codewords sorted by frequency.

Specific description

Video frames of typically 384x288, 376x280, 320x240, 192x144, 160x120 or 128x96 pixels (see figure 1) are divided into pixel blocks, typically 8x8 pixels in size (see figure 2), and also into pixel blocks, typically 2x2 pixels in size, called mini-blocks (see figure 3). In addition, the video frames are divided into Noah

regions (see figure 4), indicating how complex an area of the image is.

In one implementation, each super-block is divided into regions, each region in each super-block approximating the corresponding pixels in the original image and containing the following information:

1 Y values (typically 8 bits)

1 U value (typically 8 bits)

1 V value (typically 8 bits)

64 bits of mask specifying which YUV value to use when reconstructing this super-block.

In this implementation, each mini-block contains the following information:

2 Y values (typically 8 bits each)

1 U value (typically 8 bits)

1 V value (typically 8 bits)

4 bits of mask specifying which Y value to use when reconstructing this mini-block.

Temporal gaps

If more latency is acceptable, temporal gaps rather than spatial gaps turn out to be an efficient representation. This involves coding each changed mini-block with a codeword indicating the next time (if any) in which it changes.

Interpolation between Uniform Super-Blocks

at least includes a representation of the luminance of a block component of at least eight by eight individual pixels (super-block);

establishing a reduced number of possible luminance values for each block of pixels (typically no more than four);

encoding to derive from the words representing individual pixels further words describing blocks or groups of pixels each described as a single derived word which at least includes a representation of the luminance of a block component of typically two by two individual pixels (mini-block);

establishing a reduced number of possible luminance values for each block of pixels (typically one or two);

providing a series of changeable stored masks as a means for indicating which of the possible luminance values are to be used in determining the appropriate luminance value of each pixel for display;

comparing and evaluating the words representing corresponding portions of one frame with another frame or frames in a predetermined sequential order of the elements making up the groups to detect differences and hence changes;

identifying any of the masks which require updating to reflect such differences and choosing a fresh mask as the most appropriate to represent such differences and storing the fresh mask or masks for transmission or storage;

using context which will be available at the time of decompression to encode the masks, the changes in Y values, U values, and V values, and the spatial or temporal

gaps between changed blocks, combined with the efficient encoding scheme, to give an efficient compressed real time representation of the video;

using variable length codewords to represent the result of transitions in a way which is nearly optimal from a compression point of view, and computationally very efficient to calculate.

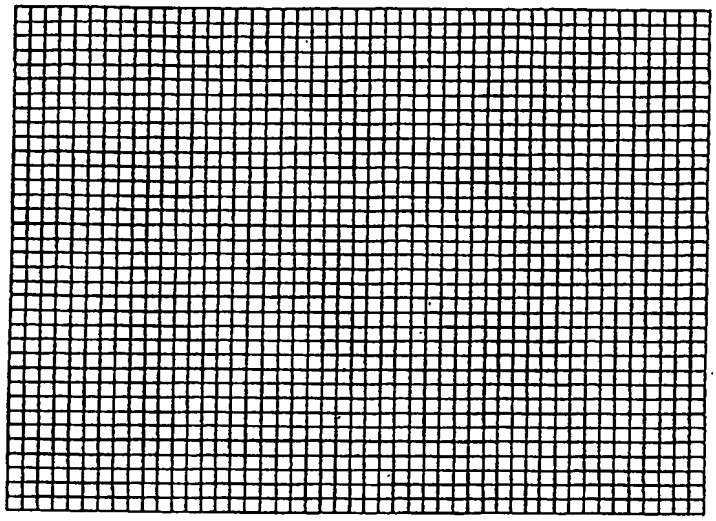


Figure 1

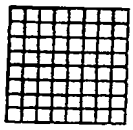


Figure 2

Figure 3

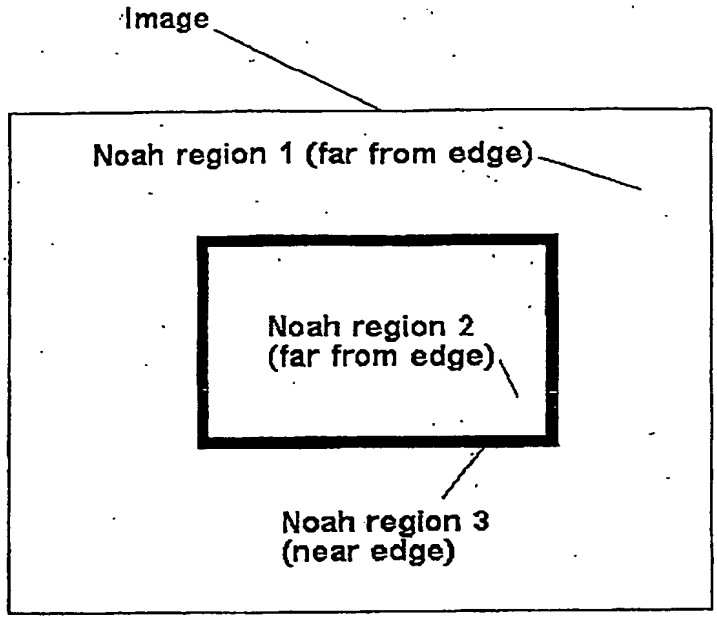


Figure 4

Global context

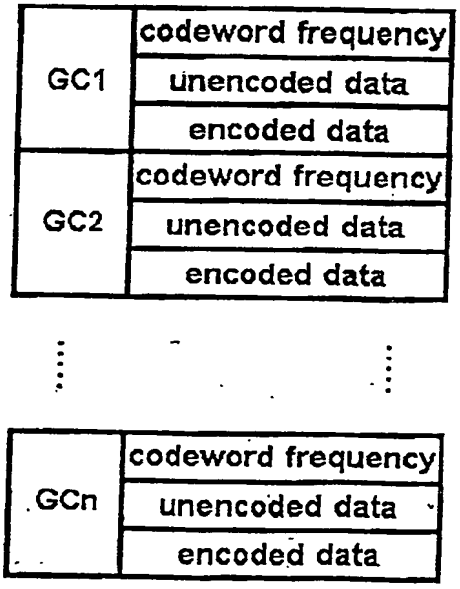


Figure 5

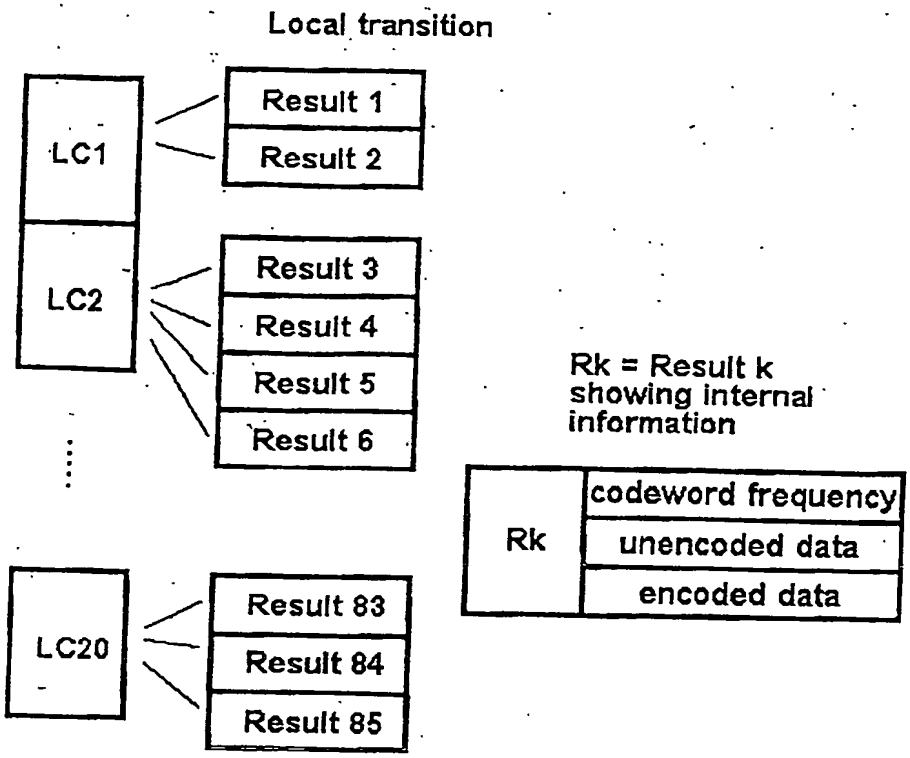


Figure 6

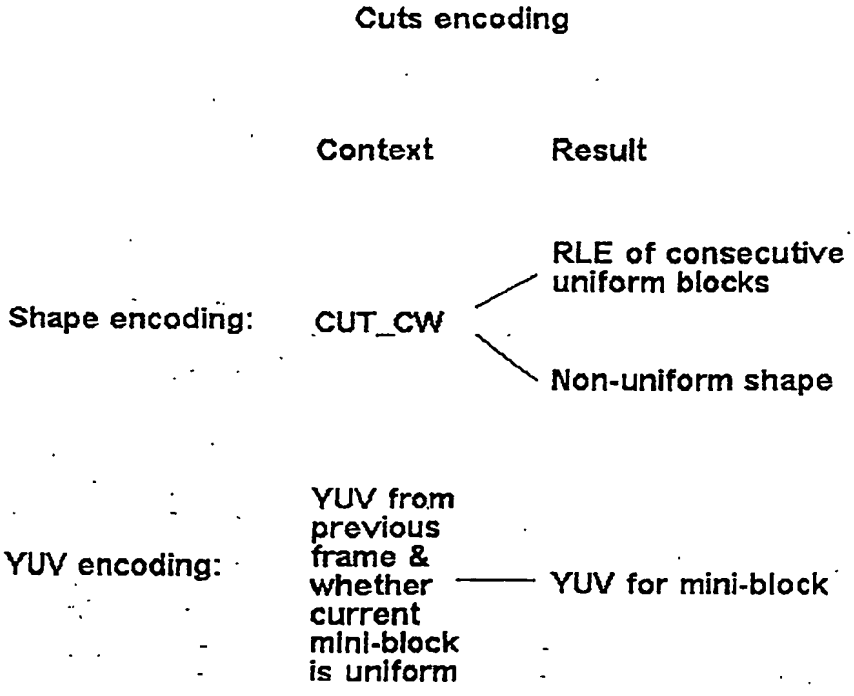


Figure 7

Delta encoding

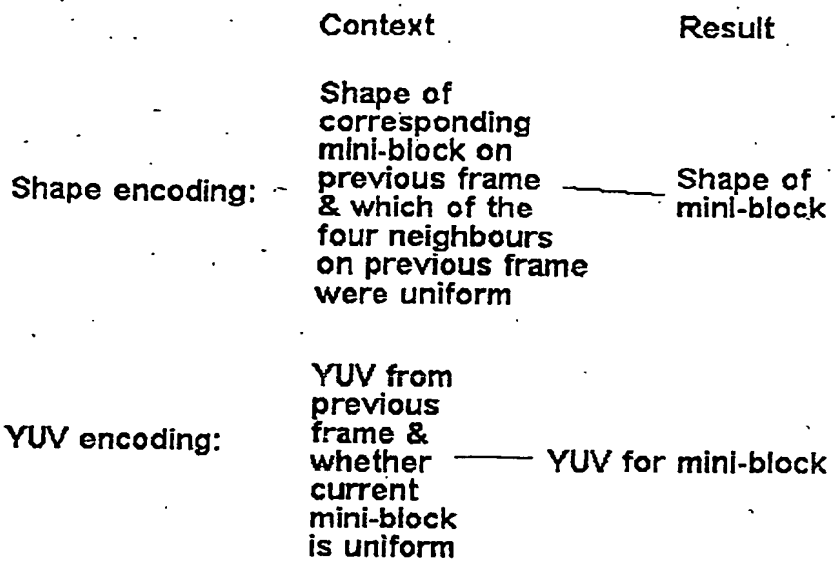


Figure 8

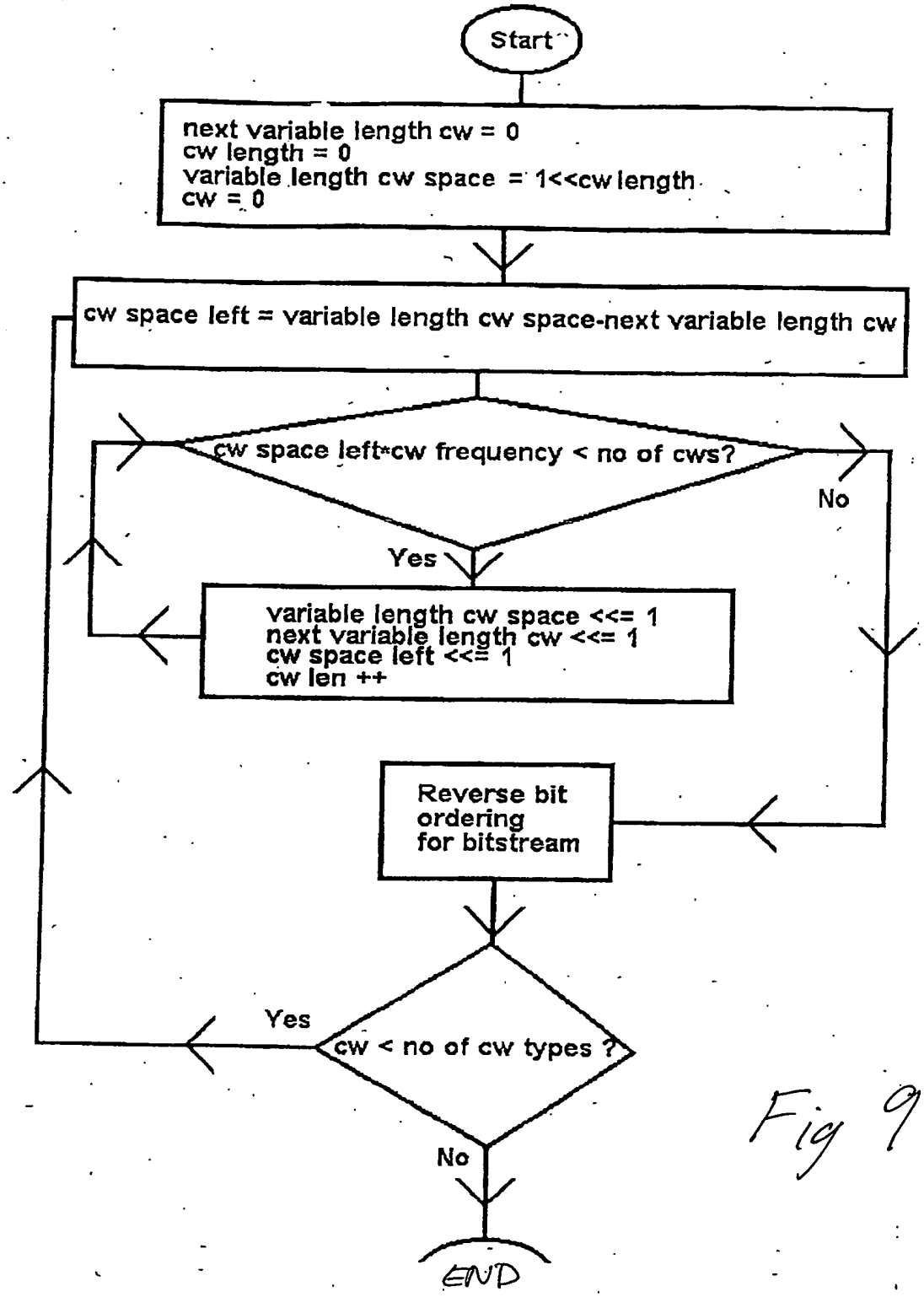

Start
next variable length cw = 0
cw length = 0
variable length cw space = 1<<cw length
cw = 0
cw space left = variable length cw space-next variable length cw
cw space left*cw frequency < no of cws?
No
Yes
variable length cw space <<= 1
next variable length cw <<= 1
cw space left <<= 1
cw len ++
Reverse bit ordering for bitstream
Yes
cw < no of cw types ?
No
END

Fig 9

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/GB04/004758

International filing date: 10 November 2004 (10.11.2004)

Document type: Certified copy of priority document

Document details: Country/Office: GB
Number: 0326201.1
Filing date: 10 November 2003 (10.11.2003)

Date of receipt at the International Bureau: 25 February 2005 (25.02.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.